



Pipelines de ETL e Machine Learning com Apache Spark

Projeto 4 Pipeline de Pré-Processamento, Validação Cruzada e Otimização em Machine Learning

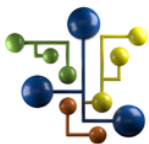


Este projeto traz o pipeline completo de pré-processamento, validação cruzada e otimização de hiperparâmetros de um modelo de Machine Learning. O pipeline é executado em ambiente distribuído no Cluster Spark, carregando os dados diretamente do sistema de armazenamento distribuído.

Os dados usados neste projeto foram extraídos do link abaixo:

<https://archive.ics.uci.edu/dataset/2/adult>

Bons estudos.



Equipe DSA

Muito Obrigado!
Continue Trilhando Uma Excelente Jornada de Aprendizagem.